Proceedings Paper:

Chua, Alton Y.K. and Banerjee, Snehasish orcid.org/0000-0001-6355-0470 (2018) Rumors and Rumor Corrections on Twitter : Studying Message Characteristics and Opinion Leadership. In: 2018 4th International Conference on Information Management (ICIM). International Conference on Information Management, 25-27 May 2018 IEEE , GBR , pp. 210-214.

https://doi.org/10.1109/INFOMAN.2018.8392837

Rumors and Rumor Corrections on Twitter: Studying Message Characteristics and Opinion Leadership

Alton Y. K. Chua
School of Communication and Information
Nanyang Technological University
Singapore
e-mail: altonchua@ntu.edu.sg

Snehasish Banerjee
The York Management School
University of York
York, UK
e-mail: snehasish.banerjee@york.ac.uk

***Abstract*—As rumors often ripple across the cyberspace, posting rumor corrections on social media can bring about social good by spreading the truth. However, rumors and rumor corrections are not easily distinguishable from one another. Therefore, this paper investigates how three message characteristics, namely, the use of emotions, clarity and credible source attribution, can predict message veracity on social media. Message veracity denotes whether a message is a rumor or a rumor correction. In addition, the paper further examines the extent to which opinion leadership moderates the relation between message characteristics and message veracity. Set against the context of the death hoax of Singapore's first Prime Minister Lee Kuan Yew in March 2015, data for this paper came from Twitter. Analysis involved binary logistic regression. All the three message characteristics predicted veracity. Rumor corrections were characterized by lower use of emotions, higher clarity, and higher credible source attribution compared with rumors. Furthermore, opinion leadership moderated the relation between the use of emotions and message veracity as well as that between credible source attribution and message veracity.**

Keywords-opinion leader, rumor, rumor correction, source credibility, Twitter

I. INTRODUCTION

Crises and emergencies often trigger rumors on social media especially when timely official information is not forthcoming. Rumors refer to unverified yet instrumentally relevant online messages that propagate fast and wide in situations of ambiguity. Despite being unverified however, rumors can be used by social media users to make sense of uncertain situations [1]. As they ripple across the cyberspace, rumors engender misunderstanding and strike panic.

Given the far-reaching ramifications, online rumormongering has been a major subject of scholarly attention in recent years. The dominant research themes include the function of rumors as collective sense-making [2], factors that have a bearing on rumor transmission [3], and users' perceptions of rumors [4]. The context of these studies run the gamut from politics [5] and natural disasters [6] to health [3] and organizational crises [7].

More recently, research has started to cast the spotlight on the use of rumor corrections to curb the effect of rumors on social media. A few nascent themes can be identified in the literature. For example, one line of investigation identifies socio-cognitive aspects including personality, motivation, cognition and social norms to explain rumor correction behaviors [8]. A second line of investigation uses machine learning algorithms to measure the efficacy of rumor corrections in the wake of a rumor outbreak [9]. A third line of investigation relies on social network analyses to study different aspects of rumor corrections such as their virality, the most common path taken, and the most active nodes in the network [10].

Despite the growing body of literature on rumors and rumor corrections, few works have explicitly focused on differences in message characteristics between the two. Expectedly, both rumors and rumor corrections contain claims of the truth. Hence, they may not be easily distinguishable. If users fail to separate rumor corrections and rumors, the purpose of the former is defeated.

In this vein, the literature on information quality suggests that message characteristics play a crucial role in helping users heuristically assess message veracity. Specifically, verified messages are seldom laced with emotions [11]. They are generally clear with minimal scope of misinterpretation [12]. In addition, they are usually underpinned by attributions to credible sources to connote their veracity [13].

Meanwhile, the concept of opinion leadership can be used to categorize social media users into two groups: opinion leaders and opinion followers [14]. The former commands a greater fan following, and is more popular in the online community than the latter. Opinion leadership has huge implications in rumor research, which often uses epidemics as a way to model and identify influential rumor spreaders [10,15]. However, in the context of rumors and rumor corrections, little is known about the interplay of message characteristics and opinion leadership.

For these reasons, this paper seeks to investigate how three message characteristics, namely, the use of emotions, clarity and credible source attribution, can predict message veracity. In this paper, message veracity denotes whether a message is a rumor or a rumor correction. The paper further examines the extent to which opinion leadership moderates the relation between message characteristics and message veracity. Data were drawn from Twitter, where rumors and rumor corrections are known to spread easily. The paper is set against the context of the death hoax of Singapore's first Prime Minister Lee Kuan Yew in March 2015.

The paper is significant on two fronts. Theoretically, it develops and empirically tests a model that predicts message veracity on Twitter based on message characteristics and users' opinion leadership. On the practical front, it teases out specific message characteristics of rumors and rumor corrections. These have the potential to help authorities as well as Twitter users craft appropriate rumor corrections.

II. METHODS

A. Data Collection

This paper studies the death hoax case of Singapore's first Prime Minister Lee Kuan Yew on Twitter, a social media platform that has been widely used in related studies [5]. Prime Minister Lee passed away on 23 March 2015. However, on 18 March 2015, a rumor reporting his death rippled through Twitter. To debunk the rumor, several news organizations posted rumor corrections, which were also shared on Twitter [16].

Tweets related to this death hoax posted on 18 March 2015 form the data for this paper. They were retrieved using the hashtags #LeeKuanYew and #LKY. The data collection process yielded 5,885 tweets altogether.

B. Data Coding

Three research assistants were recruited for data coding. They were Singaporeans, graduate students of Information Systems in a large public university in Singapore, conversant with the use of Twitter, and familiar with the death hoax case under investigation.

This paper called for two separate data coding tasks. In the first task, tweets had to be coded as rumors and rumor corrections. Those that were neither rumors nor rumor corrections would not have been meaningful to be included for analysis. In the second task, tweets that were either rumors or rumor corrections had to be further coded in terms of the three message characteristics: the use of emotions, clarity and credible source attribution.

Given the extensive coding requirement, the initial pool of tweets was reduced to a manageable volume of 2,000 entries using simple random sampling. Every tweet in the reduced pool was subjected to the first coding task. Tweets were labelled as rumors (0) if they asserted that Lee Kuan Yew had passed away. In contrast, they were labelled as rumor corrections (1) if they confirmed that he was still alive.

Two exclusion criteria were employed. First, when a tweet was deemed to be neither a rumor nor a rumor correction by any one of the three coders, it was eliminated. Second, when there was a lack of consensus among the coders about whether a tweet was a rumor or a rumor correction, it was eliminated. These criteria resulted in the exclusion of 833 tweets.

The filtered set of 1,167 tweets (2000 - 833), which contained unanimously agreed rumors and rumor corrections, was admitted for the final analysis. This assures the validity of message veracity in the corpus. In particular, there were 596 rumors and 571 rumor corrections.

This filtered set of tweets was subjected to the second coding task. For this purpose, a two-step approach was followed. In the first step, the coders jointly coded 400 randomly-selected tweets (200 rumors + 200 rumor corrections). With respect to the use of emotions, tweets were coded as 1 if they maintained emotional tone; 0 otherwise. With respect to clarity, tweets were coded as 1 if they conveyed intrinsic, contextual or representational information quality; 0 otherwise. Intrinsic information quality denotes innate clarity to foster believability and objectivity. Contextual information quality denotes the clarity of the information in the particular context at hand. It enhances relevance and timeliness of information. Finally, representational information quality measures the degree to which the information is easy to process. It improves reading ease and interpretability [17,18]. With respect to credible source attribution, tweets were coded as 1 if they made references to genuine credible sources such as media outlets and news agencies; 0 otherwise. Disagreements among the coders were resolved through discussion. This allowed establishing inter-coder agreement that was measured in terms of Cohen's Kappa (κ). Specifically, the mean pair-wise inter-coder agreement among the three coders was 0.74 for the use of emotions, 0.77 for clarity, and 0.83 for credible source attribution. A value above 0.60 for all confirmed a non-chance level of agreement. In the second step, the remaining 767 tweets (1167 - 400) were distributed uniformly among the coders to be coded independently.

Contributors of the tweets in the final dataset were dichotomized as either opinion leaders or opinion followers based on a median split of their social network size. Users' social network size was calculated as the number of followers they had on Twitter. The higher the network size of a user, the greater would be the reach of the tweets posted [19].

C. Data Analysis

Binary logistic regression was used for data analysis. It was appropriate because the dependent variable message veracity was binary—coded as 1 for rumor corrections and 0 for rumors. The three message characteristics, namely, the use of emotions, clarity and credible source attribution, comprised the independent variables.

Opinion leadership was incorporated in the model as a moderating variable. In other words, three product terms were added: use of emotions x opinion leadership, clarity x opinion leadership, and credible source attribution x opinion leadership. The direct relation from opinion leadership to message veracity was controlled. To delve deeper into the nature of the moderating relationship, follow-up binary logistic regression analyses were conducted separately for opinion leaders and opinion followers.

Prior to the analysis, the dataset was checked for the problem of multicollinearity. All values of variance inflation factor were less than the recommended threshold 10. The highest variance inflation factor value was 1.87, corresponding to the product term: credible source attribution x opinion leadership. This confirmed that the dataset was free from multicollinearity.

III. RESULTS

Table I presents the descriptive statistics of the dataset corresponding to all the three message characteristics. The use of emotions was more prevalent in rumors (55.20%) than in rumor corrections (10.51%). Clarity was higher in rumor corrections (90.02%) than in rumors (30.03%). Credible source attribution was more likely to occur in rumor corrections (73.56%) vis-à-vis rumors (6.04%).

Table II presents the detailed results of the binary logistic regression analysis. Omnibus test confirmed a significant model performance (χ^2 =827.45, df = 7, Nagelkerke R^2 = 65.90%). The control variable opinion leadership was negatively related to message veracity (B = -0.29, $p < 0.05$).

TABLE I. DESCRIPTIVE STATISTICS

	Coded as	#Rumors (%)	#Corrections (%)
Use of emotions	0	267 (44.80%)	511 (89.49%)
	1	329 (55.20%)	60 (10.51%)
Clarity	0	417 (69.97%)	57 (9.98%)
	1	179 (30.03%)	514 (90.02%)
Credible source attribution	0	560 (93.96%)	151 (26.44%)
	1	36 (6.04%)	420 (73.56%)

With respect to the message characteristics, the use of emotions was negatively related to message veracity (B = -1.31, $p < 0.001$). This meant that rumor corrections were less likely to be emotional in tone compared with rumors. Next, clarity was positively related to message veracity (B = 2.15, $p < 0.001$). Stated otherwise, rumor corrections were more likely to have clear content compared with rumors. In addition, credible source attribution was positively related to message veracity (B = 2.91, $p < 0.001$). Put differently, rumor corrections were more likely to contain credible source attributions compared with rumors.

Finally, two of the three product terms showed significant relations with the dependent variable: use of emotions x opinion leadership (B = 0.31, $p < 0.05$), and credible source attribution x opinion leadership (B = -0.46, $p < 0.05$). However, the product term clarity x opinion leadership was non-significant (B = -0.12, $p > 0.05$). In other words, opinion leadership moderated the relation between the use of emotions and message veracity as well as that between credible source attribution and message veracity.

To further validate the role of the moderator, the dataset was split based on opinion leadership and separately analyzed using binary logistic regression. The independent variables were the use of emotions and credible source attribution, which were included to delve deeper into the significant moderations.

The relation between the use of emotions and message veracity was non-significant among opinion leaders (B = -0.68, $p > 0.05$) but significantly negative among opinion followers (B = -1.98, $p < 0.001$). Thus, compared with opinion leaders, opinion followers were more likely to post non-emotional rumor corrections.

The relation between credible source attribution and message veracity was relatively weaker among opinion leaders (B = 1.99, $p < 0.001$) vis-à-vis opinion followers (B = 3.89, $p < 0.001$). Thus, compared with opinion leaders, opinion followers were more likely to attribute rumor corrections to credible sources.

TABLE II. BINARY LOGISTIC REGRESSION RESULTS

	B	SE	Exp(β)
Opinion leadership (OL)	-0.29*	0.12	0.75
Use of emotions	-1.31***	0.20	0.27
Clarity	2.15***	0.19	8.57
Credible source attribution	2.91***	0.22	18.30
Use of emotions x OL	0.31*	0.10	1.36
Clarity x OL	-0.12	0.10	0.88
Credible source attribution x OL	-0.46*	0.14	0.63

* $p < 0.05$, *** $p < 0.001$

IV. DISCUSSION AND CONCLUSION

A. *Summary of the Findings*

This paper sought to investigate how three message characteristics, namely, the use of emotions, clarity and credible source attribution, along with users' opinion leadership could help predict message veracity, which refers to whether a tweet was a rumor or a rumor correction. It was set against the context of the death hoax of Singapore's first Prime Minister Lee Kuan Yew in March 2015 on Twitter.

The results indicated that all the three message characteristics helped differentiate between rumors and rumor corrections. In particular, rumor corrections were characterized by lower use of emotions, higher clarity, and higher credible source attribution compared with rumors. Furthermore, opinion leadership moderated the relation between the use of emotions and message veracity as well as that between credible source attribution and message veracity. The binary logistic regression model validated in the paper showed promising explanatory power (Nagelkerke R^2 = 65.90%).

At a granular level, four major findings could be gleaned from the results. First, rumor corrections were less likely to be emotional in tone compared with rumors (B = -1.31, $p < 0.001$). In this vein, prior works provide evidence that rumors could be rich in emotion words in order to appear sensational [20]. After all, it is common for people to express themselves through emotionally charged messages in an ambiguous situation surrounding the death hoax of a political figure [21]. Consistent with the literature, this paper confirms the affective nature of rumors. For example, an emotional rumor tweet expressed grief stating, "*So sad at the passing of my idol #LeeKuanYew.*" On the other hand, emotions were conspicuously rare in rumor corrections. An example of such a tweet is as follows: "*#LeeKuanYew #StillAlive.*" One of the few emotional rumor corrections angrily remarked, "*Faking someone's death so fun meh? Damn lame.*" It seems that netizens who post rumor corrections are not generally driven by emotions.

Second, rumor corrections were more likely to have clear content compared with rumors ($B = 2.15$, $p < 0.001$). Rumor corrections were often very clear in articulating a course of action to the online community. This is evident from tweets such as "*Kindly do not spread rumors about Mr #LeeKuanYew*" and "*Stop saying he is dead.*" In contrast, rumors were generally ambiguous. This can be seen, for example, from the excerpt of a rumor tweet, "*He could be having his last few breath as I am writing this…*" The inherently speculative nature of rumors prevented them from being triangulated with evidences. This is why rumors generally appear more spurious and conjured vis-à-vis rumor corrections [1,17,18,20].

Third, rumor corrections were more likely to attribute to credible sources compared with rumors ($B = 2.91$, $p < 0.001$). In fact, of all the three message characteristics studied in this paper, credible source attribution had the strongest relation with message veracity. Unlike rumors, rumor corrections frequently made references to credible sources such as CNN and the Prime Minister's Office of Singapore as a way to indicate their veracity. Source credibility has long been known to be a crucial aspect of persuasive communication. Credible sources engender trustworthiness [22], believability [23], and accuracy [24] by serving as cognitive authorities. Therefore, this finding suggests a healthy trend that social media users honor the basic rule of cognitive authority while tweeting even in the wake of a social crisis [13,20].

Finally, opinion leaders did not necessarily craft more appropriate rumor corrections compared with opinion followers. The literature suggests that rumor corrections ideally should not be emotional but must contain credible source attribution [13,20]. Counter-intuitively however, compared with opinion leaders, opinion followers were more likely to post non-emotional rumor corrections ($B = 0.31$, $p < 0.05$), and to make references to credible sources ($B = -0.46$, $p < 0.05$).

With respect to the use of emotions, it seems that opinion leaders did not shy away from venting their feelings even in rumor corrections. Given their large fan following, they were bold enough to deviate from orthodox strategies in debunking rumors. Their rumor corrections were often fervent embellished with traces of anger and admonition rather than simple unimpassioned facts (e.g., "*Sadden by stupid ppl [people] spreading such sick news on #LKY*"). This can also be attributed to the political sensitivity and polarizing nature of the issue.

With respect to credible source attribution, opinion leaders did not always make references to credible sources perhaps because of their standing and reputation in the online community. In contrast, opinion followers perhaps knew that if they did not refer to credible sources in dispelling rumors, their rumor corrections would fall on deaf ears. Being less assertive, they could be keener to attribute their rumor corrections to credible sources.

While further investigation is needed to shed greater light on this new and unexpected finding, the fact that even opinion followers can sometimes post apt rumor corrections is encouraging. In fact, opinion followers seem to be doing better than opinion leaders in crafting rumor corrections with limited emotions but with credible source attribution. Building on previous works [25], this finding hints that the online community has the capability of self-correction and self-policing when presented with unverified information regardless of opinion leadership.

B. Contributions and Implications

This paper makes two major theoretical contributions. First, it demonstrates that message characteristics can help separate rumors from rumor corrections even though both contain claims of the truth and may not be easily distinguishable. This dovetails previous works that had shown the possibility to separate rumors from non-rumors [20] as well as true rumors from false rumors [25,26].

Second, this paper combined two approaches in the study of rumor corrections: one based on message characteristics, and the other based on users' opinion leadership. The robustness of the logistic regression results (Nagelkerke $R^2 = 65.90\%$) advances scholarship in the area of rumor corrections. In particular, it takes a small step toward establishing the autonomy of rumor correction research which has so far been mostly reliant on rumor research (e.g., [8,20]). This autonomy is necessary in the long run as it will eventually pave the way for a greater diversity of research focusing on different facets of rumor corrections in a variety of contexts.

The paper also has implications for practice. By identifying message characteristics of rumor corrections, it can assist organizations' public relations or communication managers when planning a rumor rebuttal for public dissemination on social media. Furthermore, the three message characteristics, all of which proved to be significant predictors of message veracity, could be used by software companies to develop classifiers to predict the veracity of online messages. Efforts can be made to automate the manual coding process as much as possible. The results of such classifiers can be used to recommend rumor corrections to users, and weed out rumors. This could be a modest step toward curbing the spread of rumors via social media. In addition, this paper recommends users—both opinion leaders and opinion followers alike—to act responsibly on social media, and craft appropriate rumor corrections.

C. Limitations and Future Research Directions

A few limitations in this paper need to be acknowledged. For one, it is limited by its choice of data sources. It analyzed tweets related to a single case. Hence, caution is advocated in generalizing its findings. Moreover, as with any research that draws data from social media, this paper failed to uncover differences in users' motivations to post rumors and rumor corrections on Twitter. It derived its findings solely based on the analyzed tweets. The link between contributors' motivation and message veracity still remains unknown.

For scholars interested in this research area, this paper identifies three future directions. First, scholars could build on the findings of this paper to predict message veracity in a variety of contexts such as celebrity gossip, organizational crisis, and disaster hoax. Second, user studies could be conducted to examine the extent to which social media users' perceptions of rumors differ from those of rumor corrections. Third, information and communication scholars could explore ways to phrase rumor corrections so that the messages are maximally effective in debunking rumors. This paper hopes that such lines of inquiry will minimize the likelihood of Internet users from being taken in easily by rumors in the long run.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education Research Grant AcRF Tier 2 (MOE2014-T2-2-020). The authors would like to thank Eddy Effendi, Wang Dezhou, and Yong Kim Fook for their help in the data collection and coding.

REFERENCES

[1] N. DiFonzo, and P. Bordia, Rumor psychology: Social and organizational approaches. Washington, WA: American Psychological Association, 2007.

[2] A. Rojecki, and S. Meraz, "Rumors and factitious informational blends: The role of the web in speculative politics," New Media & Society, vol. 18, pp. 25-43, 2016.

[3] A. Chua, and S. Banerjee, "To share or not to share: The role of epistemic belief in online health rumors," International Journal of Medical Informatics, vol. 108, pp. 36-41, 2017.

[4] J. Qi, S. Banerjee, and A. Chua, "Analyzing medical personnel's perceptions of online health rumors," International MultiConference of Engineers and Computer Scientists, pp. 457-460. Hong Kong: IAENG, 2017.

[5] J. Shin, L. Jian, K. Driscoll, and F. Bar, "Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction," New Media & Society, vol. 19, pp. 1214-1235, 2017.

[6] M. C. Allaire, "Disaster loss and social media: Can online information increase flood resilience?," Water Resources Research, vol. 52, pp. 7408-7423, 2016.

[7] A. Pal, A. Chua, and D. Goh, "Does KFC sell rat? Analysis of tweets in the wake of a rumor outbreak," Aslib Journal of Information Management, vol. 69, pp. 660-673, 2017.

[8] L. Zhao, J. Yin, and Y. Song, "An exploration of rumor combating behavior on social media in the context of social crises," Computers in Human Behavior, vol. 58, pp. 25-36, 2016.

[9] R. Tripathy, A. Bagchi, and S. Mehta, "Towards combating rumors in social networks: Models and metrics," Intelligent Data Analysis, vol. 17, pp. 149-175, 2013.

[10] K. Ji, J. Liu, and G. Xiang, "Anti-rumor dynamics and emergence of the timing threshold on complex network," Physica A: Statistical Mechanics and its Applications, vol. 411, pp. 87-94, 2014.

[11] M. T. Pham, "Emotion and rationality: A critical review and interpretation of empirical evidence," Review of General Psychology, vol. 11, pp. 155-178, 2007.

[12] A. Chua, and S. Banerjee, "A study of tweet veracity to separate rumours from counter-rumours," International Conference on Social Media & Society, pp. 4:1-8. New York, NY: ACM.

[13] S. Y. Rieh, and D. R. Danielson, "Credibility: A multidisciplinary framework," Annual Review of Information Science and Technology, vol. 41, pp. 307-364, 2007.

[14] Y. Zhao, and G. Kou, "Bounded confidence-based opinion formation for opinion leaders and opinion followers on social networks," Studies in Informatics and Control, vol. 23, pp. 153-162, 2014.

[15] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted LeaderRank," Physica A: Statistical Mechanics and its Applications, vol. 404, pp. 47-55, 2014.

[16] F. Chan, and W. Sim, "Police looking into hoax website that falsely announced death of Mr Lee Kuan Yew," The Straits Times, 19 March 2015, available at: http://www.straitstimes.com/singapore/police-looking-intohoax-website-that-falsely-announced-death-of-mr-leekuan-yew

[17] M. Ghasemaghaei, and K. Hassanein, "Online information quality and consumer satisfaction: The moderating roles of contextual factors – A meta-analysis," Information & Management, vol. 52, pp. 965-981, 2015.

[18] R. Y. Wang, and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," Journal of Management Information Systems, vol. 12, pp. 5-33, 1996.

[19] M. Cha, H. Haddadi, F. Benevenuto, and P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," International Conference on Weblogs and Social Media, pp. 10-17. Palo Alto, CA: AAAI, 2010.

[20] O. Oh, M. Agrawal, and H. R. Rao, "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises," MIS Quarterly, vol. 37, pp. 407-426, 2013.

[21] N. Smyrnaios, and P. Ratinaud, "The Charlie Hebdo attacks on Twitter: A comparative analysis of a political controversy in English and French," Social Media + Society, vol. 3, 2017, available at: http://journals.sagepub.com/doi/full/10.1177/2056305117693647

[22] P. Wilson, Second-hand knowledge: An inquiry into cognitive authority. Westport, CT: Greenwood Press, 1983.

[23] C. Wathen, and J. Burkell, "Believe it or not: Factors influencing credibility on the Web," Journal of the American Society for Information Science and Technology, vol. 53, pp. 134-144, 2002.

[24] B. Hilligoss, and S. Y. Rieh, "Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context," Information Processing & Management, vol. 44, pp. 1467-1484, 2008.

[25] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 Boston marathon bombing," iConference Proceedings, 2014, available at: https://pdfs.semanticscholar.org/60e5/5a7c6d47ed830ac6ec5eee433d1907ad78b4.pdf

[26] Z. Zhang, Z. Zhang, and H. Li, "Predictors of the authenticity of Internet health rumours," Health Information & Libraries Journal, vol. 32, pp. 195-205, 2015.